

SOP 13: CSV Data (FQ_J Creation)

Version 1

Revision History Log:

Prev. Version #	Revision Date	Updated By	Changes Made	New Version #
-	Sept 2011	Bill Johnson	Initial version	1

1.0 Summary

After the year-end FQ_E deliverable is certified, the proprietary YSI, Inc. data are converted to commonly readable comma-delimited file form and stored in a file with extension CDF. One CDF file is produced for each site/year, based on the most recent FQ_E downloaded from each site. If the final FQ_E from a site is defective, the Project Lead uses earlier FQ_Es to recover the most observations possible.

An internal Southeast Alaska Network (SEAN) augmentation program is executed to add columns to the CDF file that support data management and quality control functions, resulting in an augmented file whose name ends in CSV. The augmentation program also pulls grading scores into the CSV from certified FQ_C site visit worksheets. The Project Leader flags erroneous values and exceptions in each file. The files are packaged into on ZIP file. The Project Lead and Data Manager iterate validation cycles until the CSV product is certified. The certified product is made available on the web.

2.0 Detailed Steps

1. Project Leader: Assemble Baseline Deliverable

- Obtain the season's certified FQ_E .dat sonde output files.*
 - Browse the production web site to “Raw Sonde Data” in the freshwater water quality main page.
 - Download the correct zip file to \\INPGLBAFS03\DATA\SEAN_Data\Work_Zone\FQ\yyyy\FQ_J\, where “yyyy” is the current year of study, creating the folder if necessary.
 - Unzip the file contents into the ...FQ_J\ directory.
- Build the CDF from the correct subset of .dat files.*
 - Review the last .dat of the year for each site by visualizing it in EcoWatch proprietary software.
 - Normally this should contain all the data for the year for a single site (Figure SOP13.1), and the earlier checkpoint .dat files are of no interest.*
 - However, if it does not cover the entire period, review the monthly checkpoint files to determine if the missing data are available in one (or more) of them.*
 - If needed data for a site are for some reason split among two or more files, the Project Leader will need to extract CDFs from each site and cut and paste one*

single CDF file out of them using Excel or other software. The timestamp of each record must be used to determine exact location of splices.

4. *In the event data are missing and unrecoverable, there is no work-around. The CDF file for the year will have gaps.*

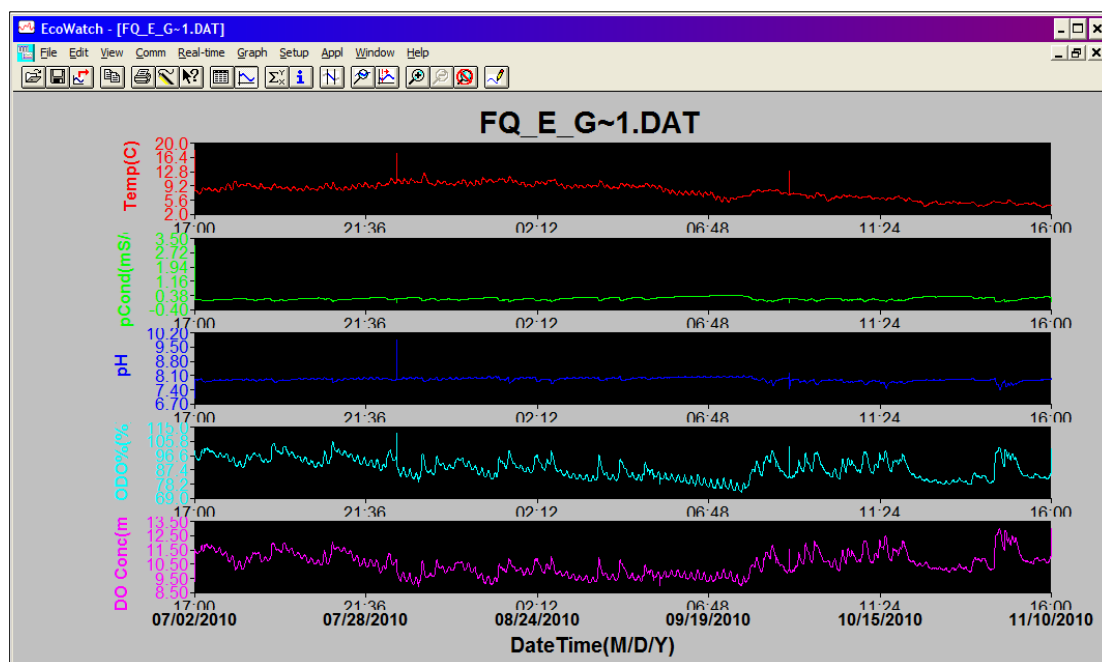


Figure SOP 13.1. EcoWatch depiction of a complete .DAT file covering the entire season. Note the characteristic temperature spikes encountered when the sonde was pulled from the water during site visits.

2. For each site, generate a CDF file using EcoWatch.
 1. *Start up the proprietary EcoWatch application available from YSI, inc.*
 2. *Do a FILE/OPEN on the most recent .dat file for the site. EcoWatch will typically read the data and draw graphs.*
 3. *Verify all needed parameters have been enabled in this file (Figure SOP13.2).*
 1. **Click on the menu sequence SETUP/PARAMETERS/ADD/REMOVE...**
 2. **In the dialog box that appears, the list labeled “Selected Parameters” must contain these values: “DateTime”, “Temp”, “SpCond”, “pH”, “ODO%”, and “ODO Conc”.**
 3. **If the selected parameters are incorrect, click on items to add or remove and press the appropriate screen button.**

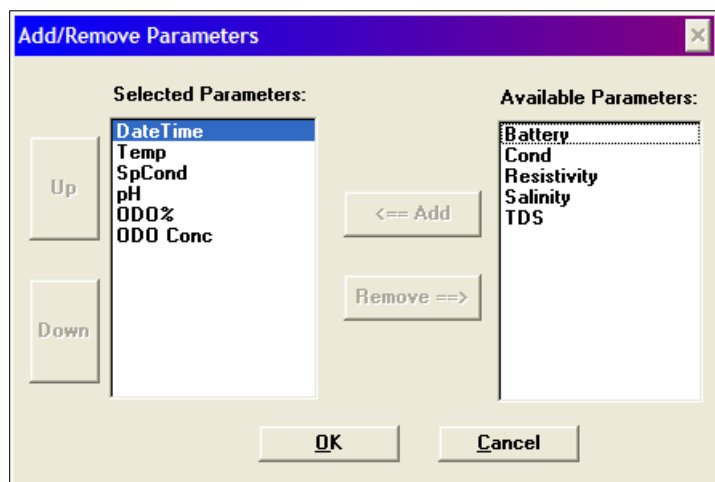


Figure SOP 13.2. Selecting the correct parameters for the CDF output file.

4. *Export the data to a comma-separated value file.*
 1. Click on the menu sequence **FILE/EXPORT/CDF/WMF...**
 2. In the **FILE EXPORT** dialog box that appears, be sure the radio button is selected for “Comma Delimited (CDF).” Also ensure the “Separate Time/Date” check box is checked on.
 3. Navigate the output to **INPGLBAFS03\DATA\SEAN_Data\Work_Zone\FQ\yyyy\FQ_J** (Figure SOP13.3).
 4. Accept and note the default filename provided. It is not practical to alter it at this stage. If this is not the first export for this year/site, then you will be asked whether to replace an older file of the same name – click **YES**.
 5. Press the **EXPORT** button. A file will be created, though no screen message announces this.
 6. Using Windows Explorer, locate the output file (by timestamp) and rename it to **FQ_J_ss_yyyy.CDF**, where **ss** is the formal site code as defined in Appendix A, and **yyyy** is the year data were recovered in.
5. *Repeat for each site.*

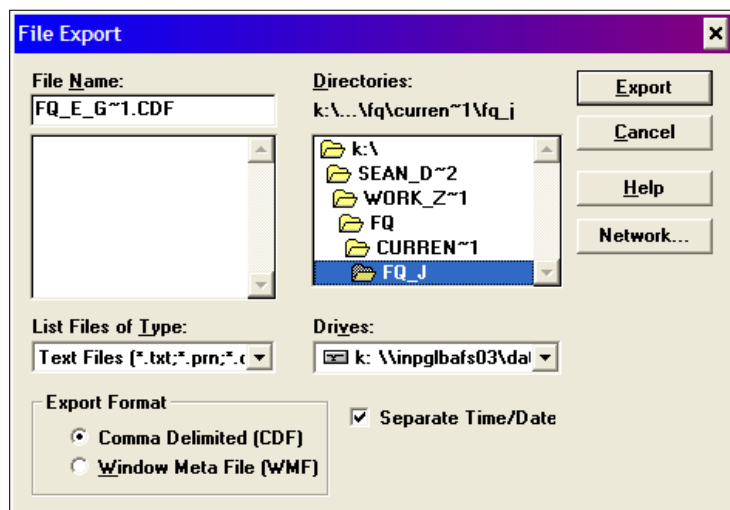


Figure SOP13.3. Specifying the comma delimited file export type and location.

2. Project Leader: Augment the CDF Content with Metadata and Grading.

1. *Invoke the FQ_J Augmenter tool on the data management website to apply a onetime adjustment, adding additional columns to each CDF, assigning data grades where possible, and saving new versions having CSV extensions.*
2. *Use the dropdown lists to specify the year on which to work. It may take a moment to verify there is a certified FQ_C available for the year.*
3. *The application will find every file created in step 1 having the standard name FQ_J_ss_yyyy.CDF.*
4. *The application will generate in that same folder output files named FQ_J_ss_yyyy.CSV, one for each site for which there is a CDF.*
5. *Fill the required data fields on the screen to reflect metadata for the file; this is available from the FQ_C site visit worksheets as well as the text of transmittal emails.*
6. *Press the button to create the augmented files.*

3. Project Leader: Flag the Data Quality for Each CSV.

1. *Review optional errors in previous validation findings of FQ_J.*
2. *Review FQ_C certified site visit worksheets online and note exceptions that indicate questionable or invalid data.*
3. *Perform any visualizations or analyses, such as time series graphing, required to detect questionable parameter readings due to events listed in Section 4 below.*
4. *Open each site's file in Excel or other software and assign Record Qualities.*
 1. *Assign each sensor's record_quality column with "0" for good when there is no question as to its correctness. This is the default set by the Augmenter.*
 2. *Mark data quality as "1" if it is not possible to ascertain the validity of a reading. This case typically occurs when the site visit and maintenance logs have been lost for an instrument and it is not known whether or not exceptions were encountered. Then enter a quality comment explaining why "1" was assigned.*
 3. *Mark sensor reading quality as "2" where appropriate to indicate questionable values, which may or may not be suitable for particular researchers' purposes. Then enter a quality comment explaining why "2" was assigned. Reasons may be noted as comments on the site visit worksheets. Typical reasons for assigning "2" include:*
 1. *Inconsistent sonde readings with no known environmental cause.*
 2. *Suspected tampering or vandalism.*
 3. *Deviation from proscribed protocol methodology.*
 4. *Mark quality as "3" where appropriate to indicate invalid values that may not be used for any purpose, as may occur when an instrument is physically damaged. Then enter a quality comment explaining why "3" was assigned. Typical reasons for "3" include:*
 1. *Failed sensor.*
 2. *Weak batteries.*
 3. *Physical sonde damage.*
 4. *Scrambled .DAT file recorded.*
 5. *Sonde washed out.*

5. For any record `_quality` that is changed from zero, its corresponding `_grade` should be deleted. `_grade` is to be present only for good records.
6. For records affected by exceptional circumstances noted on the site visit forms it may be useful to explain these in the COMMENTS column near the end of the record.
7. Resave the processed file as CSV, but leave it open.
5. **Flag erroneous values**
 1. Using expert knowledge and the criteria specified in SOP 2, determine any value of any sensor that is an erroneous value.
 2. Change the value of its corresponding erroneous value attribute to 'Y'. No other value is acceptable.
 3. Resave the correct CSV.
 4. Close the file.
6. *Repeat for each site.*

4. Project Leader: Build and Submit the Deliverable File.

1. *Use Winzip or similar software to generate a single archive file in ...\\yyyy\FQ_J\ containing all the CSV files in that folder.*
2. *Rename the deliverable to FQ_J_yyyy.ZIP, where yyyy is the calendar year represented.*
3. *Submit the candidate deliverable for validation by sending a formal email to the Data Manager specifying that FQ_J as defined in protocol FQ-2013.1 is available for review. Also specify which year the data cover.*

5. Data Manager Tasks

1. *On receipt of submission email, generate a new submission record using the tool on the data management web site. Fill in the details through the Submission Date field.*
2. *Create a subdirectory bearing the name of the assigned submission number in \\NPGLBAFS03\DATA\SEAN_Data\Staging\FQ\FQ_J\.*
3. *Copy into the subdirectory the FQ_J_yyyy.ZIP file from \\NPGLBAFS03\DATA\SEAN_Data\Work_Zone\FQ\yyyy\FQ_J\.*
4. *If this is the first submission of FQ_J for the year, update the deliverable tracking spreadsheet with its date.*
5. *Execute the automated validation process on the data management web site. The results will be automatically recorded on the submission log.*
6. *If submission fails mandatory criteria, reply with a "failure email" that includes the submission number assigned, deliverable ID, protocol ID, year covered, and the "findings.htm" file generated during validation which details the file faults.*
7. *If, instead, submission passes all mandatory criteria, reply with a "success email" that includes the submission number assigned, deliverable ID, protocol ID, year covered, the "findings.htm" file generated during validation which will detail option errors, and a request to certify the deliverable as complete.*

6. Project Leader Tasks

1. *On receipt of a failure email:*
 1. Open the underlying site CSV files with Excel and apply corrections to resolve the listed mandatory errors.

2. Consult with the Park Lead if necessary.
3. Restart the submission process from the beginning.
2. *On receipt of a success email:*
 1. Review any failed optional criteria.
 2. If the optional faults are acceptable, then reply to the Data Manager with a “certification email” stating the deliverable is certified and may be disseminated.
 3. If the faults are unacceptable, then reply with a “withdrawal email” stating the deliverable candidate is to be withdrawn. Then take remedial action to correct the optional faults and restart the submission process from the beginning.

7. Data Manager Tasks

1. *On receipt of a withdrawal email:*
 1. Mark the withdrawal in the Submission_Log’s Status column using the tool on the data management web site.
 2. Terminate processing.
2. *On receipt of a certification email:*
 1. Verify no sensitive information is in the deliverable. Products containing sensitive information cannot be disseminated. (Sensitivity is unlikely for this deliverable.)
 2. If the deliverable is judged as sensitive:
 1. *Copy the submitted file to test environment at AUXREP\FQ\FQ_J\SENSITIVE\.*
 2. *Update the Submission_Log’s Status to sensitive.*
 3. *Update the test web site with an entry for this year marked to be sensitive and closed.*
 4. *Propagate from test to production environment.*
 5. *Verify deliverable is NOT accessible from production web site.*
 6. *Update the deliverable tracking spreadsheet with the date of completion for FQ_J.*
 3. If the deliverable is judge not to be sensitive:
 1. *Copy the submitted file to test environment at AUXREP\FQ\FQ_J\.*
 2. *Propagate from test to production environment.*
 3. *DO NOT verify deliverable is accessible from production web site. Because FQ_J is expected to be replaced by a deliverable that conforms to upcoming guidance from NPS Water Resources Division, FQ_J is not to be disseminated until further notice.*
 4. *Mark the certification in the Status column in Submission_Log.*
 5. *Update the deliverable tracking spreadsheet with the dates of completion and dissemination for FQ_J.*
3. *Update the scope of the formal deliverable metadata so it includes this new date range.*